

Maier, Uwe

Wolfgang Böttcher / Hans-Georg Kotthoff (Hrsg.): Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung. Münster: Waxmann 2007 (233 S.) [...] [Sammelrezension]

Erziehungswissenschaftliche Revue (EWR) 7 (2008) 4



Quellenangabe/ Reference:

Maier, Uwe: Wolfgang Böttcher / Hans-Georg Kotthoff (Hrsg.): Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung. Münster: Waxmann 2007 (233 S.) [...] [Sammelrezension] - In: Erziehungswissenschaftliche Revue (EWR) 7 (2008) 4 - URN: urn:nbn:de:0111-pedocs-196053 - DOI: 10.25656/01:19605

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-196053>

<https://doi.org/10.25656/01:19605>

in Kooperation mit / in cooperation with:



<http://www.klinkhardt.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Erziehungswissenschaftliche Revue (EWR)

**Rezensionszeitschrift für alle Teilbereiche der
Erziehungswissenschaft
ISSN 1613-0677**

Die Rezensionen werden in die Zeitschrift mittels eines
Peer-Review-Verfahrens aufgenommen.

Weitere Informationen: <http://www.klinkhardt.de/ewr/>

Kontakt: EWR@klinkhardt.de

Bildungsstandards und Evaluation - Eine Sammelbesprechung

Wolfgang Böttcher / Hans-Georg Kotthoff (Hrsg.)

Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung

Münster: Waxmann 2007

(233 S.; ISBN 978-3-8309-1678-9 ; 24,90 EUR)

Dietrich Benner (Hrsg.)

Bildungsstandards

Chancen und Grenzen, Beispiele und Perspektiven

Paderborn: Schöningh 2007

(249 S.; ISBN 978-3-5067-6331-0 ; 29,90 EUR)

Ingmar Hosenfeld / Jana Groß Ophoff (Hrsg.)

Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht

Landau: Verlag Empirische Pädagogik 2007

(124 S.; ISBN 978-3-9373-3367-0 ; 15,90 EUR)

Das Thema Bildungsstandards ist auch mehrere Jahre nach der großen Verwunderung über nur mediokre Leistungen des deutschen Durchschnittsschülers ein zentrales Thema der erziehungswissenschaftlichen und bildungspolitischen Diskussion. Allerdings haben sich die thematischen Schwerpunkte verlagert. Während zu Beginn dieses Jahrzehnts vor allem über die Aussagekraft von Schulleistungsmessungen und die Konzeption von Standards diskutiert wurde, fragt man sich nun, wie Standards, Tests und neue Steuerungsideen tatsächlich Unterricht und Schule verbessern können.

Um genau diese Fragestellung kreisen die Beiträge der drei Herausgeberbände. Dabei zeigt die Heterogenität der Ansätze, mit welcher Vielfalt sich die erziehungswissenschaftliche Theoriebildung dieser Problematik nähern kann. Das ist beeindruckend und zeugt von einer lebendigen, kritischen Wissenschaft. Andererseits wird zu einem so populären Thema wie Bildungsstandards auch vieles geschrieben, was eigentlich unnötig ist oder in anderen Zusammenhängen besser aufgehoben wäre. Das Ergebnis der Rezension deshalb schon einmal vorneweg:

Trotz der sowieso schon bestehenden Fülle an Literatur in diesem Bereich und trotz thematischer Ausrutscher hier und da sind diese drei Bücher ein „must read“ für alle Akteure und Wissenschaftler, die mit der Umsetzung oder Erforschung von Bildungsstandards – in welcher Weise auch immer – befasst sind.

Benner: Bildungsstandards

Das erste Buch ist ein von Dietrich Benner herausgegebener Tagungsband zum Symposium *Bildungsstandards* im Jahr 2006 an der Humboldt-Universität zu Berlin. Für die Beiträge gab es anscheinend keine weitere thematische Eingrenzung und somit liegt eine Textsammlung vor, in der sich namhafte aber auch unbekannte Autoren auf einem sehr abstrakten und theoretisch anspruchsvollen Niveau grundsätzliche und manchmal auch schwer nachvollziehbare Gedanken zu Chancen und Grenzen von Bildungsstandards sowie zu deren Ausgestaltung in reflexiv-

domänenspezifischen Kompetenzmodellen machen. Das Gute an der Lektüre ist, dass die 15 Einzelbeiträge auf über 250 Seiten einen Spannungsbogen erzeugen. Man weiß als Leser nie genau, was einen als nächstes erwartet: eine abwägende Analyse, ein kritisch-konstruktiver Vorschlag zur Weiterentwicklung oder der Versuch, Abgründe der Standards aufzuzeigen. Als roter Faden hätte sich durchaus der Begriff „reflexive Bildungsstandards“ geeignet. Benner leistet in seinem Beitrag eine entsprechende Begründung und auch im Klappentext wird er erwähnt. Dabei bleibt es aber.

Der Band wird eröffnet von einem „mainstream“-Beitrag. Zur Einstimmung erzählt Olaf Köller noch einmal die Geschichte von PISA und der darauf folgenden Einführung von Bildungsstandards. Thematisch drehen sich die Ausführungen dann um die standard-analogen, einheitlichen Prüfungsanforderungen für die allgemeine Hochschulreife sowie um ländervergleichende Ergebnisse der TOSCA-Studie. Zwar ist darüber schon an anderen Stellen berichtet worden, aber Köller kann exemplarisch noch einmal schön aufzeigen, dass man mit kompetenzorientierten Tests große Schulform- und Länderunterschiede in der Leistung von Abiturienten aufzeigen kann. Daraus wird geschlussfolgert, dass mit Standards und den darauf bezogenen Testinstrumenten die Vergleichbarkeit von Schulabschlüssen geprüft und damit prinzipiell Maßnahmen zur Qualitätssteigerung angestoßen werden können. Wie diese Qualitätsmaßnahmen aussehen können und wie sie mit Standards zusammenhängen, wird den nachfolgenden Autoren überlassen. Köller geht es vor allem um das Bewusstsein für die Differenzen und um die ordentliche Messung dieser Differenzen.

Sowohl Helmut Heid als auch Jörg Ruhloff arbeiten mit kritischen Begriffsanalysen und machen in ihren Beiträgen auf Widersprüche aufmerksam, die entstehen, wenn man mit Begriffen wie Kompetenzen und Standards eine Qualitätsverbesserung des Bildungswesens betreiben möchte. Beispielsweise geht Ruhloff zunächst einmal von der Alltagssprachlichen Bedeutung von „Standard“ aus und kommt zu interessanten Vergleichen: Wiederholbare und gleichförmige Produktionsprozesse benötigen Standardisierung, um Qualität und Effizienz zu sichern. Bei der Maßanfertigung eines Anzugs, einem einmaligen und auf das Individuum bezogenen Produktionsprozess besteht die Qualität gerade darin, dass keine Orientierung an Standards möglich ist. Solche Gedanken regen zum Nachdenken an. Der Höhepunkt der Argumentation ist allerdings die exakte Herausarbeitung eines ausschließenden Verbindlichkeitsanspruchs von Bildungsstandards. Gemeinhin werden Standards mit Zielen oder Richtlinien gleichgesetzt, sozusagen als etwas Altbekanntes dargestellt. Ruhloff sieht gerade an dieser Stelle eine entscheidende Differenz. Ziele und Richtlinien implizieren die Möglichkeit des Scheiterns, des Nichterreichens des gesetzten Ziels, Standards dagegen müssen erreicht werden. Die Möglichkeit, dass Bildung auch scheitern kann, besteht nicht mehr.

Sowohl Heid als auch Ruhloff bewegen sich auf einem Terrain weitab von der empirisch-sozialwissenschaftlichen Auseinandersetzung mit Standards – sowohl sprachlich als auch inhaltlich. Wieder etwas empirischer und sozialwissenschaftlicher wird es dagegen bei Peter Zedler und Hans Merckens. Zedler vertritt die These, dass Qualitätssicherung im Bildungswesen die Voraussetzungen für einen „erziehenden Unterricht“ ausblendet und das auch noch gegen die umfangreiche Befundlage im Bereich Motivations- und Interessenforschung. Die Begründungen dieser These sind einerseits schlüssig, andererseits aber auch immer wieder sperrig und langatmig.

Einen wesentlich handfesteren Ansatzpunkt für die Kritik der bisherigen Standardisierungspraxis sowie für die pädagogische Weiterentwicklung standardbasierter Leistungsmessungen liefert auf jeden Fall Merrens in seinem gut strukturierten und verständlich geschriebenen Beitrag. Er stellt sich darin die Frage, wie Steuerung durch Standards und zentrale Leistungsmessungen überhaupt möglich sei. Rückmeldungen sind für Schule und Lehrer ja eigentlich nichts Neues. Das Problem liegt vielmehr darin, dass diese Informationen im System unterkomplex interpretiert werden, d.h. sie werden nicht im vollen Umfang zur Steuerung eingesetzt. Eine Beobachtung, die vermuten lässt, dass die neuen, sehr elaborierten Rückmeldeinformationen ein ähnliches Schicksal erleben werden. Sehr instruktiv ist an dieser Stelle der Vergleich mit Steuerungskonzepten aus anderen gesellschaftlichen Bereichen, vor allem der Industrie. Merrens arbeitet heraus, dass neuere Qualitätssicherungssysteme in der Ökonomie vor allem eine Prozesssteuerung bezwecken. Eine reine Outputsteuerung ist für Produktionsprozesse zu wenig aussagekräftig. Aber gerade diese Form der Steuerung wird momentan im Bildungswesen favorisiert. Die Frage ist somit, ob überhaupt Steuerungsinformationen für die entscheidenden Prozesse im Bildungswesen vorliegen. Auch die von PISA zu Recht angeprangerte Unfähigkeit des Schulsystems im Umgang mit heterogenen Schülergruppen wird durch eine in Standards verdeckte „Tendenz zur Homogenisierung“ (90) nicht bearbeitet.

Nach einem Exkurs von Fritz Oser zur Frage der Standardisierbarkeit von Lehrerkompetenzprofilen geht es wieder zurück zum Hauptthema. Der Herausgeber selbst klärt in seinem Beitrag die Relation von Unterricht, Wissen und Kompetenz, indem die Unterschiede zwischen Testaufgaben und didaktischen Aufgaben dargelegt werden. Hierzu vergleicht Benner auf eine interessante und instruktive Weise eine PISA-Fernsehshow-Aufgabe mit einem ähnlichen didaktischen Arrangement in Rousseaus Emile. Der Leser wird dabei in einen bildungstheoretischen Argumentationszusammenhang auf hohem Niveau verstrickt. Unter Rückgriff auf Platons Höhlengleichnis einerseits und neuzeitliche Positionen bei Descartes und Bacon andererseits, zeigt Benner, dass Wissen und Kompetenz nicht als Duale gedacht werden können. Vielmehr verlässt der Lernende den Raum der Erfahrung, die Höhle, um die übergeordneten Ideen und somit auch das eigene Nicht-Wissen zu erkennen. Gerade diese Negativität der Erfahrung hat dann eine bildende Bedeutung bei der Rückkehr in die Höhle und führt im antiken Verständnis zu einer reflexiven Kompetenz und einer Partizipationskompetenz. Leider ist dieser Artikel viel zu kurz und es bleibt zum Schluss bei stichwortartigen Andeutungen, wie diese beiden Kompetenzformen fach- und domänenspezifisch auszugestalten sind.

Anschließend leitet Benner mit einem Zwischenbericht zum DFG-Projekt *RU-Bi-Qua* (Qualitätssicherung und Bildungsstandards im evangelischen Religionsunterricht) den zweiten Teil des Bandes ein. Ausgehend vom Unbehagen, dass die Bildungspolitik vor allem an der Entwicklung von Kompetenzmodellen für „harte“ Kernfächer interessiert ist, wollte die *RU-Bi-Qua*-Forscherguppe wissen, ob auch für die sog. „weichen“ Fächer, wie z.B. Religionsunterricht, Kompetenzmodellierungen mit entsprechenden Tests möglich sind. Ähnlich wie in den Kernfächern muss allerdings auch hier ein äußerst eingeschränkter Begriff von religiöser Kompetenz in Kauf genommen werden. Der für den Religionsunterricht zentrale Bezugspunkt der Spiritualität und Glaubensstärke wurde einfach ausgeklammert. Übrig bleibt dann beispielsweise eine Skala „religiöse Deutungskompetenz“, die sehr stark der Lesekompetenz-Skala in PISA gleicht. Das Forscherteam diskutiert dieses Resultat

allerdings selbstkritisch und kann doch noch anhand einzelner Aufgabenschwierigkeiten eine domänenspezifische Komponente der Skala aufzeigen. Das lässt für weitere Entwicklungen hoffen.

Die weiteren Beiträge verzichten dann wieder auf empirische Daten und versuchen sich an begrifflichen Klärungen und kritischen Analysen bisheriger Kompetenz-Standardisierungen. Michele Borrelli macht sich Gedanken über die Aporetik als Grundform moralischer und ethischer Diskurse, jedoch ohne eine für Kompetenzmodellierungen anschlussfähige Antwort zu finden. Elk Franke vertritt die These, dass Kernfächer nicht unbedingt als Muster für Bildungsstandards in ästhetisch-expressiven Schulfächern geeignet sind und Christopher Winch warnt vor einer Herabqualifizierung beruflicher Bildung in Deutschland, wenn man sich zu sehr am Qualitätsrahmen der Europäischen Union orientiert.

Lesenswert ist vor allem eine Darstellung der chinesischen Diskussion über Bildungsstandards, neue Unterrichtskonzepte und das Verhältnis von Wissen und Können. Xu Binjan argumentiert, dass auswendig gelerntes Basiswissen und die Geschwindigkeit bei Rechenoperationen das traditionelle Fundament einer effektiven Bearbeitung mathematischer Probleme im chinesischen Unterricht war. Die Teilnahme von China-Hong Kong an PISA führte allerdings trotz sehr guter Resultate zu einer gesamtchinesischen Diskussion über Kompetenzorientierung und eine Neuausrichtung des Mathematikunterrichts. Der Beitrag mündet in ein Plädoyer für die Aufhebung der irrtümlichen Dualität zwischen Wissens- und Kompetenzorientierung und in Vorschlägen für eine sinnvolle Weiterentwicklung des überzogen anwendungsorientierten Konzepts von Mathematikunterricht.

Im dritten Teil des Bandes werden unter „Perspektiven“ noch drei Artikel zusammengefasst, die mit der Thematik mehr oder weniger zu tun haben. Eher entfernte Assoziationen zum Thema Bildungsstandards ergeben sich bei den Überlegungen zu unterrichtlichen Austausch- und Täuschungsprozessen von Roland Reichenbach und der Abhandlung über die politische Dimension der Erziehung von Philippe Foray. Bei Foray beispielsweise beschränkt sich die inhaltliche Verknüpfung zur aktuellen Standard- und Kompetenzdebatte auf den ersten Abschnitt und die drei letzten Sätze. Schon die versehentlich falsche Nummerierung verrät, dass hier einem bereits vorhandenen Manuskript vermutlich etwas angestückelt wurde.

Eine Perspektive für die Diskussion über Bildungsstandards auf einem sehr hohen, allgemeindidaktischen Niveau bietet dann doch noch der Beitrag von Klaus Prange. Er wiederholt zunächst einmal seine Vorstellungen von der Zeigestruktur der Erziehung und kommt dabei zu interessanten Anmerkungen. Beispielsweise zeigt Prange dem Leser, dass beim Prüfen eigentlich nichts anderes geprüft wird als die Zeigekompetenz des Lernenden. Ebenso kann Prange aus der Zeigestruktur der Erziehung schlussfolgern, dass die zurzeit häufig bemühte Unterscheidung zwischen Sozial-, Sach- und Methodenkompetenz nur eine theoretische ist, weil der Akt des Zeigens immer schon die drei Dimensionen in sich vereinigt. Die Pointe des Textes ist allerdings, dass durch den Gebrauch von Zeichen im Lehren und Lernen die Standardisierung bereits immer gegeben ist. Jedes Zeigen ist somit eine Anwendung, eine Art semiotische Kompetenz.

Die zusammengestellten Tagungsbeiträge sind sowohl in ihrer Form als auch in ihrem Inhalt sehr heterogen und bieten damit – je nach Leseinteresse – eine Fülle

anregender Gedanken aber auch Seiten, die man einfach überblättern kann. Empirisch darf man vor allem auf weitere Ergebnisse aus dem *RU-Bi-Qua* Projekt gespannt sein. Konzeptionell sollte sich die Debatte über die Nutzung von Evaluations- und Steuerungswissen auf jeden Fall an dem Beitrag von Hans Merckens abarbeiten. Und bildungstheoretisch geben Benner und Prange das Niveau einer Auseinandersetzung mit Bildungsstandards vor. Wesentlich spezifischer ist dagegen der Fokus des nächsten Buches.

Böttcher/Kotthoff: Schulinspektion

Bildungsstandards sind im Grunde nur eine Manifestation einer neuen Steuerungsphilosophie im öffentlichen Bereich. Parallel hierzu wird an einer Neugestaltung der Schulinspektion gearbeitet. Und wie bei Bildungsstandards finden sich außerhalb Deutschlands Modelle und Erfahrungen, die zumindest zur Kenntnis genommen werden sollten. Genau dieser Aufgabe widmen sich Wolfgang Böttcher und Hans-Georg Kotthoff mit einem Sammelband in der Reihe „Studien zur International und Interkulturell Vergleichenden Erziehungswissenschaft“. Auf 230 Seiten berichten zunächst einmal acht internationale Experten über ihre Erfahrungen mit den sehr unterschiedlichen Konzepten für Schulinspektion in England, Schweden, den Niederlanden und der Schweiz. Für eine ausgewogene Berichterstattung sorgt die Tatsache, dass immer ein Mitglied der Inspektionsagentur und ein Mitglied einer Lehrgewerkschaft befragt wurden. Das war eine sehr weise Entscheidung der Herausgeber. Zur Kontrastierung folgen dann noch Berichte über zwei Schulsysteme, die bewusst auf Schulinspektionen verzichten: USA und Finnland. Abgerundet wird der Band mit einem Blick auf die ersten Entwicklungen in drei deutschen Bundesländern.

Vorausschauend lässt sich schon einmal sagen, dass durch die doppelte Kontrastierung (verschiedene Länder und verschiedene Perspektiven) ein wirklicher Informationsgewinn entsteht. Die klaren Vorgaben der Herausgeber sorgen zudem für Übersicht und gute Lesbarkeit. Allerdings führt dies dann auch zu Beiträgen mit vielen Auflistungen und Spiegelstrichen. Kritisch anzumerken ist allenfalls, dass die Forschungsperspektive über weite Strecken fehlt. Die Herausgeber selbst sorgen zwar nachträglich für die wissenschaftliche Einordnung des Gelesenen. Aber durch die Reduktion der Perspektiven auf Inspektionsagenturen und Lehrervertreter hat das Buch immer wieder einen eher standespolitischen Charakter.

Der erste Aufsatz ist von Tim Key, einem Abteilungsleiter im „Office for Standards in Education“ (OFSTED), der englischen Schulinspektionsbehörde. Entlang des von Böttcher und Kotthoff formulierten Fragenkatalogs wird das englische Inspektionssystem kurz und verständlich dargestellt, selbstverständlich in englischer Sprache. Der Autor nimmt vor allem Bezug auf die einschneidenden Veränderungen des Schulsystems im Zuge des Education Reform Act in den 1980er Jahren. Dass diese Bildungsreform die Folge einer neoliberalen Politik war und zu einer bisher für England beispiellosen Zentralisierung eines sonst dezentral geregelten Schulwesens führte, wird allerdings nicht erwähnt. Ebenso wenig überzeugend ist dabei auch, dass mit eigenen Daten die Qualität des eigenen Inspektionssystems legitimiert wird.

Der zweite englische Beitrag zu OFSTED von Robert Cheesman, einem Funktionär der nationalen Lehrgewerkschaft in England, ist dagegen *wesentlich kritischer*. Beispielsweise führt die Ankündigung einer OFSTED-Inspektion an Schulen zu

außergewöhnlichen Aktivitäten mit zum Teil extremer Arbeitsbelastung für Lehrkräfte. Und auch die Veröffentlichung der Inspektionsberichte kann je nach Inhalt fatale Folgen für Schulen haben: Eltern schicken ihre Kinder an andere Schulen und qualifizierte Lehrer wandern ab.

Ähnlich informativ ist das ebenfalls englischsprachige Berichts-Double zur Schulinspektion in Schweden. Die schwedische Evaluationsfachfrau Gunilla Olson gibt einen profunden Überblick zur Evaluation im schwedischen Schulwesen und lokalisiert die Ursprünge der schwedischen Schulinspektion vor allem in Strategien der Dezentralisierung des Schulsystems in den 1990er Jahren. Die Qualität der Schulinspektion wurde über Verfahren der formativen Prozessevaluation ständig weiterentwickelt. Befunde zu „harten“ Wirkungseffekten liegen aber auch in Schweden noch nicht vor. Zumindest kann Olson nur von einer Befragungsstudie des schwedischen Finanzministeriums berichten. Naturgemäß kritischer ist die Sichtweise der schwedischen Lehrgewerkschaft. Diese setzt sich vor allem für eine Stärkung der internen Evaluation und eine Stärkung der Qualitätskontrollen auf lokaler Ebene ein.

Aus Sicht der interkantonalen Arbeitsgemeinschaft für Schulevaluation berichten Brägger, Kramis und Teuteberg über die Schulaufsichtsreform in der Schweiz am Beispiel der Kantone Luzern und Thurgau. Erwähnenswert ist beispielsweise die klare organisatorische Trennung von Schulevaluation und Schulaufsicht in zwei Abteilungen, aber unter dem Dach eines Amtes. Damit versucht man den beiden spannungsgeladenen Funktionen der Rechenschaftslegung und Qualitätsentwicklung gerecht zu werden. Schon allein aufgrund der kulturellen Nähe ist dieser Beitrag eine Fundgrube für bundesdeutsche Chefinspektoren.

Anton Strittmatter repräsentiert die Schweizer Lehrerperspektive. Sein Beitrag ist im Vergleich zum eher deskriptiv-technokratischen Stil der vorausgehenden Autoren deutlich kritischer und immer wieder sehr polemisch. Strittmatter kritisiert zunächst einmal die Literatur der letzten Jahrzehnte und den Begriffswildwuchs im Bereich Evaluation. Es sei viel geschrieben und wenig erklärt worden: „Das liegt wohl daran, dass das Evaluationsgeschäft im Grunde eine ziemlich banale Angelegenheit ist, wenn man mal einige allgemeine erkenntnistheoretische Grundlagen bzw. das Wesen von Feedbackprozessen begriffen hat“ (94). Der Leser erfährt an dieser Stelle allerdings nichts über diese Grundlagen. Dafür weist Strittmatter sehr deutlich auf das Forschungsdefizit im Bereich Schulevaluation hin. Es gibt keine Studien, die der Reliabilität von Urteilen einer Schulinspektion nachgehen und auch die Frage nach der Nutzung von Evaluationswissen wird immer nur einseitig formuliert: Was müssen Schulen tun, um die Ergebnisse besser umzusetzen?

Der eingeschobene Exkurs zur Tiefenpsychologie des Kontrolleifers ist zwar köstlich zu lesen, trägt aber wenig zur Versachlichung der Diskussion bei. Auf jeden Fall ist Strittmatters Beitrag anregender zu lesen, als die teilweise semantisch glattgeschliffenen Beiträgen zuvor. Gegen Ende des Beitrags wird es sogar pastoral, wenn über „Hoffnungen“ geredet wird: „Es gibt zwar namentlich in der Schweiz wenig, aber doch ein klein wenig Hoffnung, dass Schulevaluation je einmal einen Hofnarren-Rechtsstatus erhalten könnte, wie ihn Her Majesty's Inspectors in England und das Niederländische Inspektorat mal hatten, bevor sie gouvernemental domestiziert wurden. Schulen würden einer externen Evaluation eher trauen [...], wenn sie die Erfahrung machen würden, dass die gewonnenen Erkenntnisse [...]

auch zu Kritik an der behördlichen Bildungspolitik [...] führen würde“ (109).

Die niederländischen Berichte sind dann wieder ganz im üblichen Evaluations-Jargon und schließen die internationale Länderberichterstattung zur Schulinspektion ab. Die nächsten drei Kapitel besprechen Formen der Rechenschaftslegung in zwei Ländern, die bewusst auf Schulinspektionen verzichten: USA und Finnland. Das spannende an dieser Auswahl ist, dass die Unterschiede kaum größer sein könnten. Und nach der Lektüre vieler Konzepte ohne rechte empirische Fundierung stellt sich dem Leser sowieso die Frage, ob denn der ganze Aufwand lohnt. Die amerikanische Variante des Verzichts auf Schulinspektionen wäre allerdings ein schlechter Tausch. Der Beitrag von Doran über „test-based accountability“ in den USA deutet die eigentlichen Probleme nur an. Das von Bush unterzeichnete „No-Child-Left-Behind“ (NCLB) Gesetz von 2002 wird zwar als nicht zweckmäßig für eine interne Prozessevaluation von Schule beschrieben, aber es wird so getan, als könnte man mit gewissen Additionen daraus ein gutes System der Rechenschaftslegung und Schulentwicklung machen. Dabei wird nur in zwei oder drei Nebensätzen erwähnt, dass bisher auch unerwünschte Nebenwirkungen festgestellt wurden. Dass diese unerwünschten Nebenwirkungen mittlerweile empirisch sehr gut abgesichert sind und kritische Soziologen und Erziehungswissenschaftler in den USA darüber ausführlich berichten, wird bei Doran nicht erwähnt.

Das geschieht dann zumindest ansatzweise im zweiten USA-Bericht von Hess und Petrilli. Obwohl dieser Beitrag wesentlich deutlicher ausfällt und populäre NCLB-Opponenten zitiert werden (Kohn, Nodding, Elmore), wird auch hier auf die gesamtgesellschaftlichen Zusammenhänge einer solchen Gesetzgebung nicht genug aufmerksam gemacht. Problematisch, und vor dem Hintergrund der US-Tradition auch ungewöhnlich, ist vor allem die zentralstaatliche Verordnung eines Indikatorensystems mit eingebautem Erfolgszwang, das eingeführt wurde, um neoliberalen Vorstellungen von Schulwahlfreiheit und Wettbewerb auch im Erziehungssystem Geltung zu verschaffen.

Im Anschluss geht es weiter ins gelobte Finnland. Als Leser muss man natürlich aufpassen, dass man das Berichtete vor dem Hintergrund der PISA-Exzellenz dieses Landes nicht allzu rasch glorifiziert. Dennoch ist schon erstaunlich, wenn in einem Buch, das über die Einführung oder Neuordnung der Schulinspektion in Deutschland informieren soll, folgendes steht: „Die Abschaffung der schulischen Inspektion durch das Zentralamt für Unterrichtswesen (...) ging grundsätzlich von den Schulinspektoren selbst aus, welche die Meinung vertraten, dass Schulinspektion einmal zu wenig brauchbare Informationen über den realen Zustand des Schulwesens lieferte und zum zweiten keinen wesentlichen Beitrag zur Qualitätsentwicklung der Schulen leistete“ (166). Im weiteren Verlauf des Beitrags erfährt man dann noch mehr: Vertrauen ist die zentrale Grundlage für Qualitätssicherung im finnischen Schulwesen. Schon deswegen ist dieser Beitrag direkt nach den beiden USA-Artikeln recht gut platziert worden und sollte unbedingt gelesen werden.

In weiteren Aufsätzen werden erste Ansätze einer neu formierten Schulinspektion in drei Bundesländern vorgestellt: NRW, Bayern und Hessen. Alle drei Berichte geben einen hinreichenden Überblick über die bildungspolitischen Zielsetzungen, die Funktionen und den jeweiligen Qualitätsrahmen der externen Evaluation. Von ersten Erfahrungen kann berichtet werden, allerdings liegen auch hier noch keine

systematischen Befunde zur Akzeptanz oder Wirksamkeit vor. Trotz semantischer Differenzen, die auch schon Strittmatter für die Schweiz beklagt hat, überwiegen die gemeinsamen Grundmerkmale: Kopplung externer und interner Evaluation, Nutzung multipler Daten und Kopplung von Qualitätssicherung und Schulentwicklung, so gut das eben geht. In diesen Berichten mangelt es auf jeden Fall nicht an Schaubildern, die perfekte Evaluationsprozesse suggerieren. Was fehlt, ist die empirische Überprüfung der institutionellen Lerneffekte durch die Rückspielung von Evaluationswissen.

Böttcher und Kotthoff schließen den Kreis und fassen im letzten Beitrag die „Gelingensbedingungen einer qualitätsoptimierenden Schulinspektion“ noch einmal zusammen. Das klingt alles plausibel und Wohl dem Schulsystem, dessen Steuerleute sich konsequent daran orientieren. Allerdings sollte man auch kritisch fragen, ob die von Böttcher und Kotthoff formulierten Ansprüche nicht zu ambitioniert sind und zu überzogenen Kontroll- und Rettungsphantasien von Bildungsverwaltern in einem neu definierten Steuerungsparadigma führen (siehe Strittmatter). Noch schlimmer wäre es allerdings – und da muss man Böttcher und Kotthoff zustimmen – wenn nur die Rhetorik des neuen Paradigmas kopiert wird und den Schulen keine Freiheit gewährt wird. Dann ist auch eine neue Form der Schulinspektion nichts weiter als eine „Verdoppelung der Bürokratie“ (226).

Hosenfeld/Groß Ophoff: Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht

Abschließend soll noch das von Ingmar Hosenfeld und Jana Groß Ophoff herausgegebene Themenheft „Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht“ der Zeitschrift „Empirische Pädagogik“ diskutiert werden. Was bei Böttchers und Kotthoffs Synopse internationaler Schulinspektionserfahrungen fehlt, ist in den nachfolgend vorgestellten Beiträgen reichlich vorhanden: empirische Befunde zur Nutzung von Evaluationsdaten für die Verbesserung des Unterrichts. Und zwar geht es um ganz spezielle Evaluationsdaten: Rückmeldungen an Lehrkräfte in den bundesweit bekannten und wissenschaftlich anerkannten Vergleichsarbeitsprojekten VERA, Lernstand NRW und Kompetenztest Thüringen sowie dem Forschungsprojekt BeLesen. Eine direkte Verbindung zum Benner-Band ergibt sich über den bereits erwähnten Beitrag von Hans Merkens. Er deutet konzeptionell an, in welche Richtung sich Vergleichsarbeiten und Rückmeldestudien weiterentwickeln sollten.

Im Vergleich zu den beiden zuvor besprochenen Herausgeberbänden sind natürlich andere Voraussetzungen geben. Die Zeitschrift Empirische Pädagogik ist eine begutachtete Zeitschrift mit hohen Qualitätsansprüchen an die theoretische Fundierung und die Forschungsmethodik der Berichte. Alle Beiträge des Themenheftes sind somit stark fokussiert und forschungsmethodisch auf einem hohen Niveau. Den theoretischen Rahmen geben die beiden Herausgeber, Ingmar Hosenfeld und Jana Groß Ophoff vor, indem sie das bereits an vielen Stellen publizierte Zyklusmodell schulischer Innovation von Helmke und Hosenfeld noch einmal darstellen und erklären. Allerdings wird das sonst etwas trivial anmutende Modell nun zu Recht als „Heuristik“ bezeichnet. Ebenso erweitern die beiden Herausgeber den theoretischen Rahmen von Rückmeldestudien durch die Bezugnahme auf lernpsychologische Feedbackmodelle.

Julia Schneewind berichtet von einer Lehrerbefragung zum Umgang mit Leistungsrückmeldungen im Projekt *BeLesen*. Der Beitrag ist in dreierlei Hinsicht interessant und wichtig. Erstens erweitert sie das Zyklusmodell von Hosenfeld und Helmke um eine wichtige Komponente: Die Momente, die dem Prozess von Rückmeldungen und Rezeption vorausgehen, d.h. die Zielsetzung, Konstruktion und Implementation der Vergleichsarbeit. Zweitens: *BeLesen* ist keine Vergleichsarbeit mit dem Ziel testbasierter Rechenschaftslegung. Es geht um gezielte Lernstandsdiagnostik in einem speziellen Kompetenzbereich und die halbjährliche Wiederholung der Tests signalisiert Lehrkräften ganz deutlich den formativen Charakter der Evaluation. Eigentlich müssten die Lehrkräfte diese Form der Unterstützung ihrer Arbeit willkommen heißen und nutzen, aber Schneewind muss selbst in diesem Setting auf einige Rezeptionsmissverständnisse und Nutzungshindernisse hinweisen. Und das macht sie sehr überzeugend mithilfe qualitativer Daten (Punkt 3), die bisher viel zu wenig genutzt wurden, um das Rezeptionsverhalten an Schulen nachzuzeichnen.

In den drei Berichten aus den bekannten Vergleichsarbeitsprojekten VERA, Lernstand NRW und Kompetenztest werden jeweils noch einmal die Testkonzepte vorgestellt sowie Teilaspekte der mitlaufenden Rezeptionsstudien vertieft. Sowohl forschungsmethodisch als auch theoretisch hoch interessant ist der VERA-Beitrag. Groß Ophoff, Hosenfeld und Koch nutzen die online-Befragungsdaten von VERA, um verschiedene Rezeptionstypen unterscheiden zu können. Dies gelingt und die Typen können durchaus plausibel erklärt werden. Allerdings fragt man sich hier natürlich, ob die feinen Differenzlinien einer Clusteranalyse in der Realität tatsächlich mit substanziellen Personendifferenzen korrespondieren.

Sowohl der Beitrag von Kühle und Peek zur Rezeption der nordrhein-westfälischen Lernstandserhebungen in der Sekundarstufe I als auch die Thüringer Rezeptionsstudie von Nachtigall und Jantowski geben wichtige Einblicke in den zwar nicht immer zielkonformen, jedoch vorhandenen Umgang mit Rückmeldedaten an Schulen. Respektabel ist dabei die durchaus kritische Betrachtung der Test- und Rückmeldesysteme, die ja von den Rezeptionsforschern selbst mit entwickelt wurden. Vor allem Kühle und Peek scheuen sich nicht, die Funktionsüberfrachtung der NRW Lernstandserhebungen für die unklare Nutzung der Daten auf Schulebene mit verantwortlich zu machen.

Alle drei Rezeptionsstudien müssen sich allerdings auch folgende forschungsmethodische Kritik gefallen lassen. Es wurden bisher noch keine qualitativen Daten erhoben, um die doch recht optimistisch anmutenden Nutzungsergebnisse zu validieren. Ebenso fehlt es an echten längsschnittlichen Analysen. Schulentwicklung ist ein langwieriger Prozess und datenbasierte Schulentwicklung wird sich graduell ausbilden. Ein drittes Problem betrifft die Aggregationsebene der Daten über das schulinterne Rezeptions- und Nutzungsverhalten. Es werden Aussagen auf Lehrerebene gemacht. Ziel der datenbasierten Schul- und Unterrichtsentwicklung ist allerdings die Fachkonferenzebene und die Einzelschulebene. Mehrebenenanalytisch auswertbare Datensätze wären deshalb ein nächster Forschungsschritt.

Ein weiterer Beitrag in diesem Themenheft beschäftigt sich mit Informationssystemen im Bildungswesen. Ausgehend von Forschungen zur Mensch-Maschine-Kommunikation und Informationssystemen für Manager, gehen Breiter und Stauke

der Frage nach, wie ein Rückmeldesystem für zentrale Leistungsstudien konzipiert sein müsste, um den verschiedenen Nutzergruppen eine optimale Datenaufbereitung als Grundlage für pädagogische Entscheidungen liefern zu können. Die Parallelen zu Forschungsarbeiten über Managementinformationssysteme sind sehr aufschlussreich. Bereits in den 1970er Jahren zeigten Studien, dass Manager für ihre Entscheidungen nur einen kleinen Teil der aufwändig gesammelten und aufbereiteten Daten nutzen. Teilweise tun sie dies auch nur, um nachträglich ihre Entscheidungen zu legitimieren. Ansonsten ist der Beitrag sehr informationstechnologisch und wiederholt noch einmal, was man in der internationalen Literatur über „school performance feedback systems“ schon längere Zeit diskutiert.

Informationstechnologische Finessen und weiter optimierte Rückmeldeverfahren sollten allerdings nicht darüber hinweg täuschen, dass die Rückmeldedaten zunächst einmal für Lehrkräfte inhaltlich relevant und zuverlässig sein sollten. Diese Frage bearbeitet dann Karl-Heinz Arnold im abschließenden Beitrag des Themenheftes. Er überlegt sich, welchen didaktischen Informationsgehalt eine Vergleichsarbeitsrückmeldung überhaupt haben kann. Sein Fazit ist eine indirekte, aber doch deutliche Kritik an den moderat optimistischen Ergebnissen bisheriger Rezeptionsstudien: „Vieles spricht derzeit noch für die Vermutung, dass in Deutschland die Wertschätzung von testbasierten Schulleistungsevaluationen eher ein allgemeines Einstellungsmerkmal darstellt denn eine schulpädagogisch begründbare Nutzungspraxis kennzeichnet“ (455). Insgesamt ist das Themenheft mit Gewinn zu lesen, es informiert sowohl sachlich als auch kritisch über den aktuellen Stand in der Rezeptionsforschung.

Fazit

Sowohl der theorielastige Benner-Band als auch die international vergleichenden Berichte über Schulinspektionen als auch das empirielastige Themenheft der Empirischen Pädagogik zeigen, dass noch vieles zu klären ist, wenn Bildungsstandards zusammen mit neuen Modellen der Rechenschaftslegung tatsächlich ein pädagogischer Fortschritt sein sollen. Die drei Publikationen demonstrieren allerdings auch eindrücklich, dass sich Fachleute aus unterschiedlichen Bereichen mit der Sache kritisch und konstruktiv auseinander setzen. Daran lässt sich anknüpfen.

Uwe Maier (Schwäbisch Gmünd)

Uwe Maier: Rezension von: Böttcher, Wolfgang / Kotthoff, Hans-Georg (Hg.): Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung. Münster: Waxmann 2007. In: EWR 7 (2008), Nr. 4 (Veröffentlicht am 06.08.2008), URL: <http://www.klinkhardt.de/ewr/978383091678.html>